# Towards Perceptual Soundscape Characterization Using Event Detection Algorithms

**Félix Gontier, Pierre Aumond, Mathieu Lagrange, Catherine Lavandier, Jean-Francois Petiot**

LS2N, CNRS, École Centrale de Nantes

**Abstract**

Assessing properties about specific sound sources is important to characterize better the perception of urban sound environments. In order to produce perceptually motivated noise maps, we argue that it is possible to consider the data produced by acoustic sensor networks to gather information about sources of interest and predict their perceptual attributes.

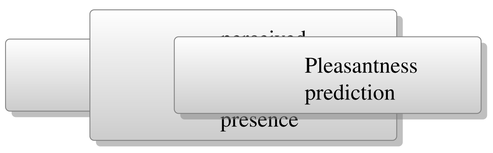
To validate this important assumption, this paper reports on a perceptual test on simulated sound scenes for which both perceptual and acoustic source properties are known. Results show that it is indeed feasible to predict perceptual source-specific quantities of interest from recordings, leading to the introduction of two predictors of perceptual judgments from acoustic data. The use of those predictors in the new task of automatic soundscape characterization is finally discussed.

## Perceptual soundscape characterization

The deterioration of **sound quality in urban** environments is an **increasing concern**. In this context, the perception of a soundscape by city dwellers is an essential part of its characterization. Perceptual spaces are constructed that associate the soundscape to subjective parameters. Specifically, the dimension of pleasantness has already been extensively studied through perceptual experiments [1]. It is consistently modeled as a function of both the sound scene's overall characteristics (loudness) and source-specific contents, typically the time of presence:

$$P = aL + \sum_s b_s T_{s,p} + c \tag{1}$$

Although previous studies focus on subjective data, the resulting models hint at the possibility of predicting pleasantness from acoustical indicators.




The corresponding task could thus be formulated as a three-level processing chain. At the first level event detection and source separation algorithms are used to determine the contents of the scene in terms of acoustical properties and source activity. The second level is the estimation of perceived time of presence for each source, for which the relevant time scale is longer. Finally, pleasantness is predicted using the available perceptual models.

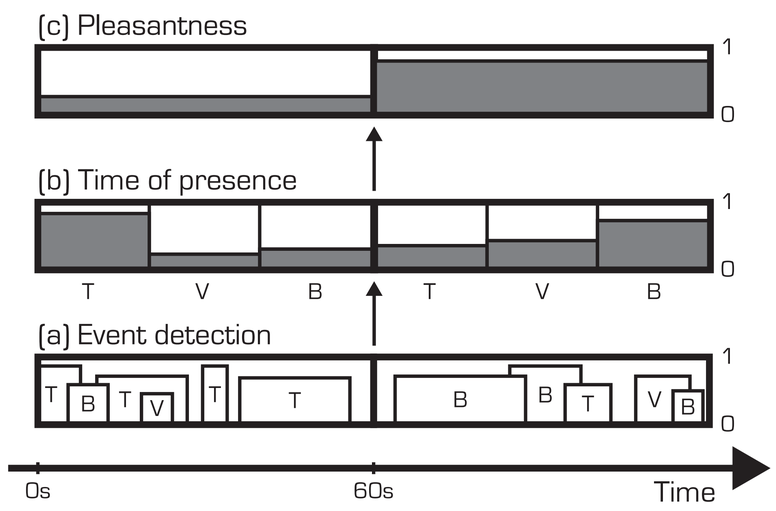



**Figure 1:** The three suggested levels of metrics to predict soundscape pleasantness. Traffic, voices and birds are considered as having the most influence on pleasantness in past studies.

## Contributions

The main contributions of this work are as follows:

- A **perceptual validation** of the feasibility of estimating the perceived source activity from acoustical data,
- The proposition of source-specific **physical indicators** in order to achieve this prediction.

## Perceptual validation

A perceptual experiment is conducted to ensure that the prediction of source-specific perceived time of presence from acoustical data is possible. A corpus of 9 simulated scenes from [3] is studied in which 5 different sources are active (traffic, birds, horns, voices, footsteps). The questions include 4 general perceptual dimensions and the source-specific time of presence and sound level. 28 students from École Centrale de Nantes performed the test. The study generates a perceptual space coherent with the litterature [2], indicating the relevance of the considered corpus.

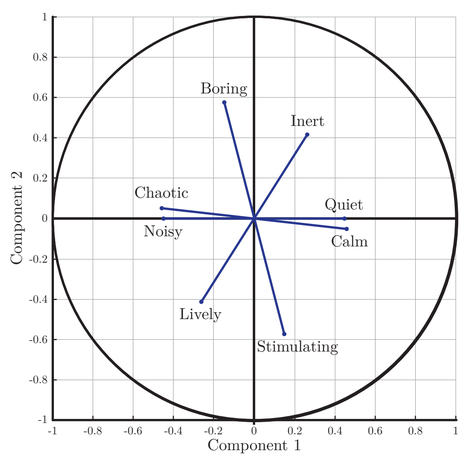



**Figure 2:** Perceptual space of the study.

## Physical indicators

Global sound levels $L_{50}$, $L_{A50}$ and $L_{50,1kHz}$ are considered as they represent well the scene's perceived overall loudness [1]. As simulated scenes are studied the ground truth source contributions are available. The time of presence and an emergence metric ($L_{10,source} - L_{90,global}$) are computed per source from the separated audio file.

To account for masking effects, two additional indicators are proposed that respectively consider overall and frequency-wise source emergence over time frames to approximate the corresponding perceived time of presence:

$$T_s(\alpha) = \frac{1}{N_t}\sum_{t=1}^{N_t} \mathbb{1}_{L_s(t)-L_b(t)>\alpha} \tag{2}$$

$$T_s(\alpha,\beta) = \frac{1}{N_t}\sum_{t=1}^{N_t} \mathbb{1}\left[\frac{\sum_{f=1}^{N_f}\Delta_s(t,f)\mathbb{1}_{\Delta_s(t,f)>\alpha}}{\sum_{f=1}^{N_f}\mathbb{1}_{\Delta_s(t,f)>\alpha}} > \beta\right], \Delta_s(t,f) = L_s(t,f) - L_b(t,f) \tag{3}$$

The optimization of parameters $\alpha$ and $\beta$ is done via grid search on the considered corpus due to the lack of other available perceptual data.

## Results

Globally computed sound levels represent very well the perceived overall loudness ($r > 0.95$, $p < 0.01$). The ground truth source-specific time of presence and emergence achieve good results for most sources. However they fail to describe traffic activity, which is present in the background of most scenes but can be masked by more emergent sources. The use of metrics considering each source independantly from one another is thus not sufficient.

The proposed binary masking models solve this problem as an active source can be considered unheard in the mix. They correlate consistently well with the perceived time of presence, and successfully discriminate between sources such as traffic and birds, which are commonly correlated in real-life conditions.

| Phys./Perc. | $L_{T,p}$ | $T_{T,p}$ | $L_{B,p}$ | $T_{B,p}$ | $L_{H,p}$ | $T_{H,p}$ | $L_{V,p}$ | $T_{V,p}$ | $L_{F,p}$ | $T_{F,p}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $T_T$ | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS |
| $L_T$ | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS |
| $T_B$ | 0.71* | 0.75* | NS | NS | NS | NS | NS | NS | NS | NS |
| $L_B$ | -0.84** | -0.83** | 0.91** | 0.82** | NS | NS | NS | NS | NS | NS |
| $T_H$ | NS | NS | NS | NS | NS | 0.84** | NS | NS | NS | NS |
| $L_H$ | NS | NS | NS | NS | 0.98** | 0.78* | NS | NS | NS | NS |
| $T_V$ | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS |
| $L_V$ | NS | NS | NS | NS | NS | NS | 0.84** | 0.88** | NS | NS |
| $T_F$ | NS | NS | NS | NS | NS | NS | NS | NS | 0.9** | 0.68* |
| $L_F$ | NS | NS | NS | NS | NS | NS | -0.69* | -0.78* | 0.92** | NS |
| $T_T(\alpha)$ | 0.90** | 0.94** | NS | NS | NS | NS | NS | NS | NS | NS |
| $T_T(\alpha,\beta)$ | 0.88** | 0.92** | NS | NS | NS | NS | NS | NS | NS | NS |
| $T_B(\alpha)$ | NS | NS | 0.95** | 0.97** | NS | NS | NS | NS | NS | NS |
| $T_B(\alpha,\beta)$ | NS | NS | 0.95** | 0.97** | NS | NS | NS | NS | NS | NS |
| $T_H(\alpha)$ | NS | NS | NS | NS | NS | 0.83** | NS | NS | NS | NS |
| $T_H(\alpha,\beta)$ | NS | NS | NS | NS | 0.73* | 0.88** | NS | NS | NS | NS |
| $T_V(\alpha)$ | NS | NS | NS | NS | NS | NS | 0.79* | 0.83** | NS | NS |
| $T_V(\alpha,\beta)$ | NS | NS | NS | NS | NS | NS | 0.75* | 0.79* | NS | NS |
| $T_F(\alpha)$ | NS | NS | NS | NS | NS | NS | NS | -0.71* | 0.87** | NS |
| $T_F(\alpha,\beta)$ | NS | NS | NS | NS | NS | NS | NS | NS | 0.90** | 0.70* |

**Table 1:** Pearson correlation coefficients between perceptual parameters and physical indicators at the scene level (n=9). *: $p < 0.05$, **: $p < 0.01$, non-significant correlations ($p > 0.05$) are noted NS.

## Conclusions

A binary masking model is proposed that relies on the time of presence and emergence of sources. It is shown to correlate well with perceptual time of presence equivalents. Pleasantness prediction in urban soundscapes task can thus be formulated as a detection and classification task. Furthermore, low time resolutions compared to traditional event detection algorithms are sufficient to capture the information at the perceptual level.

## Forthcoming Research

Our future work includes conducting a refined perceptual experiment with the following considerations:

- A richer corpus should be studied, that is not limited to real-life conditions but includes more diverse source contributions,
- State-of-the-art masking models should be compared to the proposed indicators,
- Quantities of interest should be predicted from recordings through detection or source separation algorithms.

The final objective is the formulation of a complete experimental protocol for pleasantness prediction with respect to the experimental results.

## References

[1] P. Aumond, A. Can, B. De Coensel, D. Botteldooren, C. Ribeiro, and C. Lavandier. Modeling soundscape pleasantness using perceptive assessments and acoustic measurements along paths in urban context. *Acta Acust. unit. Acust.*, 103:430–443, 2017.

[2] O. Axelsson, M.E. Nilsson, and B. Berglund. A principal components model of soundscape perception. *J. Ac. Soc. Am.*, 128:2836, 2010.

[3] J.R. Gloaguen, A. Can, M. Lagrange, and J.F. Petiot. Creation of a corpus of realistic urban sound scenes with controlled acoustic properties. In *Proceedings of Meetings on Acoustics*, 2017.

## Acknowledgements

The authors would like to acknowledge support for this project from ANR project Cense (grant ANR-16-CE22-0012).